

BEST AVAILABLE COPY

ARC920010085US1  
10/042,367In the Claims:

1. (amended herein) A method of determining parameter combinations for automated web crawler access to World Wide Web content that is accessible based on parameters resulting from real user interactions with a World Wide Web site, said method comprising:

maintaining at least one log file containing user queries resulting from previous real user

HTML interactions with said World Wide Web site;

analyzing said log file to determine parameter combinations and to generate synthetic  
queries for input to said web crawler, said web crawler using said input for automated access to  
said World Wide Web content.

Formatted: Indent: First line: 0"

Deleted: at least one set of parameters

Deleted: for

Formatted: Indent: First line: 0.5"

2. (amended herein) A method of determining parameter combinations for automated access to World Wide Web content that is accessible based on parameters resulting from real user interactions with a World Wide Web site, as per claim 1, said user queries comprising entries,  
said analyzing step further comprising

ranking entries according to their frequency of occurrence;

for a set of entries resulting from unlimited text entries, excluding entries ranked

below a predetermined number; and

wherein said synthetic queries are determined by producing combinations of entries from each set  
of entries.

Deleted: wherein said parameters are  
entries in HTML forms,

Deleted: in each set of entries

Deleted: each

Deleted: parameter combinations

3. (amended herein) A method of determining parameter combinations for automated access to World Wide Web content that is accessible based on parameters resulting from real user interactions with a World Wide Web site, as per claim 2, wherein said synthetic queries are  
determined by producing all combinations of entries from each set of entries.

Deleted: parameter combinations

ARC920010085US1  
10/042,367

4. (original) A method of determining parameter combinations for automated access to World Wide Web content that is accessible based on parameters resulting from real user interactions with a World Wide Web site, as per claim 2, wherein entries resulting from limited text entries and unlimited text entries have stop words removed and remaining words stemmed.

5. (original) A method of determining parameter combinations for automated access to World Wide Web content that is accessible based on parameters resulting from real user interactions with a World Wide Web site, as per claim 1, wherein said log file is maintained by a proxy server that logs communications between a client and a Web server resulting from real user accesses to said World Wide Web content.

6. (cancel) A method of determining parameter combinations for automated access to World Wide Web content that is accessible based on parameters resulting from real user interactions with a World Wide Web site, as per claim 1, wherein said content is automatically accessed using said parameter combinations.

Deleted: original

7. (amended herein) A method of increasing web crawler penetration of Web databases accessible via HTML forms, said method comprising:

reviewing previous real user form input data ;

identifying possible HTML form input data for said Web crawler from said previous real user form input data by synthesis of entries for any of: predefined sets, limited text entries or unlimited text entries; and

Page 5 of 12

ARC920010085051  
10/042,367

providing said identified form input data to said Web crawler during an instantiation of automated access to said Web databases by said Web crawler.

8. (currently amended) A method of increasing web crawler penetration of Web databases accessible via HTML forms, as per claim 7, wherein said previous form input data are maintained in a log file.

9. (original) A method of increasing web crawler penetration of Web databases accessible via HTML forms, as per claim 8, wherein said log file is maintained by a proxy server.

10. (original) A method of increasing web crawler penetration of Web databases accessible via HTML forms, as per claim 7, wherein said synthesis comprises:

ranking any entries for predetermined sets;

ranking any entries for limited text entries;

ranking any entries for unlimited text entries;

excluding entries for unlimited text entries ranked below a predetermined number;

and

pairing entries from each set of ranked entries.

11. (original) A method of increasing web crawler penetration of Web databases accessible via HTML forms, as per claim 10, wherein said synthesis further comprises:

removing stop words and stemming remaining words for entries resulting from limited text entries and unlimited text entries.

Page 6 of 12

ARC920010085US1  
10/042,367

12. (cancel) A method of determining entries for input items of an HTML form for automated accesses to content contained in a Web database behind said HTML form, said method comprising:

maintaining a log of real user entries for said input items;

analyzing said log to determine entry combinations for said input items.

13. (cancel) A method of determining entries for input items of an HTML form for automated accesses to content contained in a Web database behind said HTML form, as per claim 12, wherein said log file contains at least one set of entries, said analyzing step further comprising

ranking entries in each set of entries according to their frequency of occurrence;

for each set of entries resulting from unlimited text entries, excluding entries ranked below a predetermined number; and

wherein said automated parameter combinations are determined by producing combinations of entries from each set of entries.

14. (cancel) A method of determining entries for input items of an HTML form for automated accesses to content contained in a Web database behind said HTML form, as per claim 13, wherein said parameter combinations are determined by producing all combinations of entries from each set of entries.

15. (cancel) A method of determining entries for input items of an HTML form for automated accesses to content contained in a Web database behind said HTML form, as per claim 13, wherein entries resulting from limited text entries and unlimited text entries have stop words removed and remaining words stemmed.

Page 7 of 12

ARC920010085051  
10/042,367

16. (cancel) A method of determining entries for input items of an HTML form for automated accesses to content contained in a Web database behind said HTML form, as per claim 12, wherein said log file is maintained by a proxy server that logs communications between a client and a Web server resulting from real user accesses to said World Wide Web content.

17. (amended herein) A method of emulating real user access to World Wide Web content dynamically accessible via an HTML form, said method comprising:

maintaining a log containing real user entries into each input item of said HTML form;

ranking entries for each input item according to their frequency of occurrence;

for each unlimited text entry input item, excluding entries ranked below a predetermined number;

determining combinations of entries from each set of entries; and

automatically accessing said content using said combinations of entries as HTML input for a webcrawler.

18. (original) A method of emulating real user access to World Wide Web content dynamically accessible via an HTML form, as per claim 17, wherein entries resulting from limited text entries and unlimited text entries have stop words removed and remaining words stemmed.

19. (original) A method of emulating real user access to World Wide Web content dynamically accessible via an HTML form, as per claim 17, wherein said log file is maintained by a proxy

Page 8 of 12

ARC920010085US1  
10/042,367

server that logs communications between a client and a Web server resulting from real user accesses to said World Wide Web content.

20. (cancel) An article of manufacture comprising a computer usable medium having computer readable program code embed therein to determine parameter combinations for automated access to World Wide Web content that is accessible based on parameters resulting from user interactions with a World Wide Web site, said computer readable program code comprising:

computer readable program code for maintaining at least one log file representative of real user interactions with said World Wide Web site;

computer readable program code for analyzing said log file to determine parameter combinations for automated access to said World Wide Web content.

21. (cancel) An article of manufacture comprising a computer usable medium having computer readable program code embed therein to determine parameter combinations for automated access to World Wide Web content that is accessible based on parameters resulting from user interactions with a World Wide Web site, as per claim 20, wherein said parameters are entries in HTML forms, said computer readable program code for analyzing further comprising

computer readable program code for ranking entries in each set of entries according to their frequency of occurrence; and

computer readable program code for each set of entries resulting from unlimited text entries, excluding entries ranked below a predetermined number; and

wherein said parameter combinations are determined by producing combinations of entries from each set of entries.

Page 9 of 17

ARC920010085051  
10/042,367

22. (cancel) An article of manufacture comprising a computer usable medium having computer readable program code embed therein to determine parameter combinations for automated access to World Wide Web content that is accessible based on parameters resulting from user interactions with a World Wide Web site, as per claim 21, wherein said parameter combinations are determined by producing all combinations of entries from each set of entries.

23. (cancel) An article of manufacture comprising a computer usable medium having computer readable program code embed therein to determine parameter combinations for automated access to World Wide Web content that is accessible based on parameters resulting from user interactions with a World Wide Web site, as per claim 21, wherein entries resulting from limited text entries and unlimited text entries have stop words removed and remaining words stemmed.

24. (cancel) An article of manufacture comprising a computer usable medium having computer readable program code embed therein to determine parameter combinations for automated access to World Wide Web content that is accessible based on parameters resulting from user interactions with a World Wide Web site, as per claim 20, wherein said log file is maintained by a proxy server that logs communications between a client and a Web server resulting from real user accesses to said World Wide Web content.

25. (cancel) An article of manufacture comprising a computer usable medium having computer readable program code embed therein to determine parameter combinations for automated access to World Wide Web content that is accessible based on parameters resulting

Page 10 of 12

ARC920010085051  
10/042.367

from user interactions with a World Wide Web site, as per claim 20, wherein said content is automatically access using said parameter combinations.

Page 11 of 12



ARCP20010085US1  
10/042,367

As this amendment has been timely filed within the set period of response, no petition for extension of time or associated fee is required. However, the Commissioner is hereby authorized to charge any deficiencies in the fees provided to Deposit Account No. 09-0441.

If it is felt that an interview would expedite prosecution of this application, please do not hesitate to contact applicants' representative at the below number.

Respectfully submitted,



Randy W. Lacasse  
Registration No. 34,368

1725 Duke Street  
Suite 650  
Alexandria, Virginia 22314  
(703) 838-7683  
March 14, 2005

Page 12 of 12

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☒ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☒ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**